ON CHOOSING PARAMETERS SET FOR ECA HEIGHT OPTIMIZATION

Alexander Dokukin

Abstract: *A problem of choosing an appropriate set of parameters for estimates calculating algorithms height optimization is considered. The task arises for example in AVO-polynomial recognition method that depends on height optimization of individual ECA items. Initially some subset of general ECA family was used for such an optimization according to the simplicity of the resulting algorithms. In the present article it is shown that on some general conditions the defined ECA subfamily is optimal meaning that the optimal solution in more general one will belong to the said subfamily.*

Keywords: *estimates calculating algorithm, optimization, algebraic approach, pattern recognition.*

ACM Classification Keywords: *I.5 Pattern Recognition — I.5.0 General.*

Introduction

Optimization of estimates calculating algorithms (ECA) height corresponding to some set of training objects is core task of AVO-polynomial recognition method. Here ECA is a general parametrized family of recognition algorithms invented by Yu. I. Zhuravlev in 1970s [Zhuravlev, 1977b] and studied further on by him and his students. AVO-polynomial [Dokukin, 2009] is one possible application of the said family to practical recognition tasks developed by the author who is also one of Zhuravlev's students. In that particular method a subset of the initial ECA family was used for optimization of so called algorithm's height. The choosing of the subset was dictated by simplicity reasons at the time and the question of its validity remained open. In the present article it will be shown that the restrictions follow with necessity from the method design.

The first part of the article will be devoted to recalling some ECA background including basic definitions and statements: standard form of recognition task, estimates calculating algorithm and so on. ECA height optimization task as well as general AVO-polynomial method training will be described also.

In the second part the same optimization processes will be considered in a general ECA family. It will be shown that on some general conditions the optimal solution will belong to the simple subset.

Finally the nature of those conditions will be discussed giving leads to some further ECA studies.

Definitions

The standard recognition task [Zhuravlev, 1977a] is stated as follows. Let us have a training sample $\{S_1, \dots, S_{m+q}\}$, described by vectors of some nature $S_i = (a_{i1}, \dots, a_{in})$. The sample is split into l classes $K_1, \dots, K_l$ that can overlap in general case. The training sample classification that is vectors $\boldsymbol{\alpha_i} = (\alpha_{i1}, \dots, \alpha_{il})$ is known. Here α_{ij} is the value of "$S_i \in K_j$" predicate. It is required to construct an algorithm A that can calculate classification of a new object S.

Estimates calculating algorithm (ECA or AVO in Russian) refers to a general family of recognition algorithms described by Yu. I. Zhuravlev in 1970s [Zhuravlev, 1977b]. First, it is supposed that all the features are real numbers. A vector $(\varepsilon_1, \dots, \varepsilon_n)$ of non-negative ε-thresholds is used to define a proximity function of two objects $B(S_u, S_v)$

over supporting set $\omega \subset \{1,\dots,n\}$:

$$P_\omega(S_u,S_v)=\begin{cases}1, \; |a_{ui}-a_{vi}|\le\varepsilon_i, \; i\in\omega,\\ 0, \; \text{otherwise.}\end{cases}$$

Second, object estimates for different classes are introduced:

$$\Gamma_j(S)=\sum_{\omega\in\Omega_A} p_\omega \sum_{S_i\in\tilde{K}_j}\gamma(S_i)P_\omega(S,S_i)\;. \tag{1}$$

Here $\Gamma_j(S)$ is an estimate of an object's S belonging to a K_j, $\tilde{K}_j=\{S_1,\dots,S_m\}\cap K_j$ and $C\tilde{K}_j=\{S_1,\dots,S_m\}\setminus K_j$, p_ω is weight of the supporting set ω and $\gamma(S_i)$ is weight of training object S_i.

Finally, an estimates calculating operator $B(S)$ calculates a vector $(\Gamma_1(S),\dots,\Gamma_l(S))$ of estimates of a vector for each class and a decision rule C is applied converting real-value class estimates to a final decision. In the described case an object is assigned to a class of maximal estimate.

$$C(x_1,\dots,x_l)=\arg\max_{i=1,\dots,l} x_i\;.$$

The AVO-polynomial [Dokukin, 2009] method represents a polynomial over ECAs as it is supposed by its name. Algebraic operations are applied to the class estimates before the decision rule. That is the essence of the algebraic approach to recognition [Zhuravlev, 1977a]. Also in AVO-polynomial a small subset of the family is used, which corresponds to a case of a single supporting set and equal weights of training objects.

$$\Gamma_j(S)=\sum_{S_i\in\tilde{K}_j} P_{\{1,\dots,n\}}(S,S_i)\;. \tag{2}$$

This subfamily will be called the reduced one later on.

This particular polynomial is build according to the following procedure. A set of objects $\{S^1,\dots,S^q\}$ is extracted from the training sample which will be named a reference sample. As opposed to a testing one, the reference sample is used in training too, but has a different role. Each member S^t of the reference set in combination with all the remaining training objects is used for constructing one item of the polynomial B_t. The following formula describes the polynomial

$$B(S)=\sum_{S^t\in\{S^1,\dots,S^q\}} D(S)B_t(S)\;.$$

Here B_t is an ECA achieved through training and $D(S)$ is another ECA multiplier penalizing remoteness from S_t, which make the construction a second degree polynomial.

To achieve the B_t an auxiliary task is considered. For each remaining training object $S_i=(a_{i1},\dots,a_{in})$ and the $S^t=(b_{t1},\dots,b_{tn})$ a new one is constructed $S=(|a_{i1}-bt1|,\dots,|a_{in}-b_{tn}|)$. The object is assigned to a class 1 if both ancestors belong to a same class and to a class 0 otherwise. Then the optimal hyper-parallelepiped R (rectangle for short) is searched maximizing the difference between class 1 objects number and class 0 ones (ECA height).

This process of selecting a single reference object for construction of a single polynomial item is the key condition that makes the reduced ECA family optimal.

Optimization parameters

Let's consider polynomial construction in a most general way.

$$B(S)=\sum_{\sigma^t\in\{S^1,\dots,S^q\}} a_t B_t^{b_t}(S)\;.$$

Here $\sigma^t \subset \{S^1, \ldots, S^q\}$ is some subset of reference objects for which a polynomial item $B_{\sigma^t}(S)$ is constructed. $B_t(S)$ is short for $B_{\sigma^t}(S)$, $a_t = a_{\sigma^t}$ and $b_t = b_{\sigma^t}$ are coefficient and degree of item $B_t(S)$ in the polynomial.
Let's consider now a single $\sigma^t \subset \{S^1, \ldots, S^q\}$ and find an optimal ECA $B_t(S)$ in general ECA family (1).

Theorem 1. *If every object of σ^t belongs to a same class K_h then there exists a singular optimal ECA $B_t(S)$ such that $C(B(S)) = h$ for all S.*

Indeed, since the weights $\gamma(S_i)$ of training objects are variable, let's set them according to a formula:

$$\gamma(S_i) = \begin{cases} 1, \; S_i \in K_h, \\ 0, \; \text{otherwise}, \end{cases}$$

and set the ε-thresholds to fit every reference object:

$$\varepsilon_j = \max_{S_u, S_v \in \{S_1, \ldots, S_{m+q}\}} |a_{uj} - a_{vj}| + 1 \; .$$

The described algorithm will be optimal, since it gives maximum possible estimates of each reference object for its class, and gives zero estimated for every other class so the height is maximum. The algorithm is also singular, since its result is constant in considered area (which can be extended by increasing ε-thresholds). The statement is proved.

There is an evident consequence to the proved statement. It makes unreasonable using both weights of objects and ε-thresholds in height optimization in case σ_t consists of a single object. Moreover, it is unreasonable for every functional which achieves its maximum at the same point.

Let's now fix objects weights but still consider a sum of different supporting sets.

$$\Gamma_j(S) = \sum_{\omega \in \Omega_A} p_\omega \sum_{S_i \in \tilde{K}_j} P_\omega(S, S_i) \; . \tag{3}$$

Theorem 2. *Let $B_t(S)$ be an optimal algorithm from the ECA family described in (3). Then there is an algorithm $B'_t(S)$ from reduced family (2) such that $h(B_t) \leq h(B'_t)$, there $h(t)$ is algorithm's height.*

Before proving the theorem let's prove the following lemma.

Lemma 1.

$$h\left(\sum_k p_k B_k\right) = \sum_k p_k h(B_k) \; .$$

Algorithm's height by definition [Dokukin, 2009] is difference between minimal estimation of correct pairs (object, class), i. e. pair in which the object belongs to the class and a maximum estimation of incorrect pairs

$$h(B) = \min_{\{i,j | S^i \in K_j\}} \Gamma_j(S^i) - \max_{\{u,v | S^u \notin K_v\}} \Gamma_v(S^u) \; .$$

Consequently

$$h\left(\sum_k p_k B_k\right) = \min_{\{i,j|S^i \in K_j\}} \sum_k p_k \Gamma^k_j(S^i) - \max_{\{u,v|S^u \notin K_v\}} \sum_k p_k \Gamma^k_v(S^u) =$$
$$= \sum_k p_k \left(\min_{\{i,j|S^i \in K_j\}} \Gamma^k_j(S^i) - \max_{\{u,v|S^u \notin K_v\}} \Gamma^k_v(S^u) \right) = \sum_k p_k h(B_k) \; . \tag{4}$$

The lemma is proved, so let's prove the theorem.

Let's consider the optimal algorithm as a sum of single set ones

$$B_t(S) = \sum_{\omega \in \Omega_A} p_\omega B_t^\omega(S) \ .$$

Let's consider now any of its items $B_t^\omega(S)$ corresponding to $\omega \subset \{1, \ldots, n\}$. It is evident that by setting any threshold ε_i not belonging to ω, $i \in \{1, \ldots, n\} \setminus \omega$ to a great enough value we can construct $B_t^{\omega\prime}(S)$ belonging to the reduced family (2). Let now B_t' be an optimal algorithm from that family. Due to the obvious condition $\sum_{\omega \in \Omega_A} p(\omega) = 1$ and the Lemma 1 the following inequations hold

$$h\left(\sum_{\omega \in \Omega_A} p_\omega B_t^\omega(S)\right) = \sum_{\omega \in \Omega_A} p_\omega h(B_t^{\omega\prime}(S)) \leq \sum_{\omega \in \Omega_A} p_\omega h(B_t'(S)) = h(B_t'(S))$$

and the theorem is proved.

Discussion

It is now proved that if we use a single reference object for constructing an item of optimal height for the recognition polynomial it is sufficient to use the reduced family of ECA (2). In a similar fashion the same can be made for some of the other target functionals like those information value estimations used in [Dokukin, 2013] depending on whether the respective lemma can be proved.

On the other hand ECA applications are not content with the use of single reference object. Historically its applications started with the opposite approach. A single optimal ECA was searched for in accordance to the whole reference set. There are some great implementations of the approach important and used in present days like the one described in [Ryazanov, 1976]. In that algorithm ε-thresholds are fixed and optimization is performed over objects' weights with number of errors as a target functional.

There are also some peculiar uses of ECA family. For example the one aimed at recognition tasks with categorical features [Dyakonov, 2014]. It completely lacks ε-thresholds in its definition. Logical regularities based methods [Ryazanov, 2007] are also distant relatives of ECA family with their own set of parameters. For example, there is no apparent division of objects to reference and training ones. The whole set of precedents is considered as a whole, but a number of algorithms is trained in accordance to some central objects like in AVO-polynomial.

Thus, there is no clear answer as to which parameters family to use in practical ECA applications. Should any complications of well known cases provide any improvement is yet to be answered. The current work limits those possible complications to a certain degree if we consider an ECA polynomial as a target.

First of all, if we want to involve objects' weights in optimization the reference set for any target algorithm should contain objects of multiple classes as it is shown in Theorem 1. Thus, either the training vs reference division should be rearranged for each item or inevitably thresholds should be involved in optimization along with objects' weights.

Moreover, involving different supporting sets is also quite useless (see Theorem 2). It can be useful only if it is not required of the algorithm to be optimal. For example, if we use not only optimal item but some of those close to it by their functional value. Such a back-off can be useful for a better coverage of the feature space at a no cost, since it does not decelerate optimization algorithm.

Thus, the further study of classical, i. e. based on threshold similarity function, ECA optimization in recognition tasks is tightly connected to construction of reference sets. Should it be random selection or some spatial considerations is yet to be analyzed. It is clear though that such reference sets should contain objects of different classes. ECA optimization in those complicated condition is to be studied also as well as algebraic correction of the resulting algorithms set.

Acknowledgements

The paper is published with partial support by the project ITHEA XXI of the ITHEA ISS (www.ithea.org) and the ADUIS (www.aduis.com.ua) as well as Russian Foundation for Basic Research projects No. 14-01-90019 and 14-07-00965.

Bibliography

[Zhuravlev, 1977a] Zhuravlev Yu. I. Correct algebras over sets of incorrect (heuristic) algorithms I (in Russian) // Cybernetics. — 1977. — No. 4. — Pp. 14–21.

[Zhuravlev, 1977b] Zhuravlev Yu. I. Correct algebras over sets of incorrect (heuristic) algorithms II (in Russian) // Cybernetics. — 1977. — No. 6. — Pp. 21–27.

[Ryazanov, 2007] Ryazanov V. V. Logical regularities in pattern recognition (parametric approach) // Computational Mathematics and Mathematical Physics. — 2007. — Vol. 47, No. 10. — Pp. 1720–1735.

[Dokukin, 2009] Dokukin A. A. "AVO-Polynom" Recognition Algorithm // Supplement to International Journal "Information Technologies and Knowledge". — 2009. — Vol. 3, Book 8 "Classification, Forecasting, Data Mining". — Pp. 65–68.

[Dokukin, 2013] Dokukin A. A. Use of Information Value in AVO-Polynomial Method Training // International Journal "Information Models & Analyses". —2013. —Vol. 2, No. 2. —Pp. 123–126.

[Ryazanov, 1976] Ryazanov V. V. Optimization of algorithms for calculating estimates with respect to parameters characterizing the representability of the standard rows // USSR Computational Mathematics and Mathematical Physics. — 1976. — Vol. 16, No. 6. — Pp. 173–184.

[Dyakonov, 2014] D'yakonov A. G. Methods for solving classification tasks with categorical features (in Russian) // Applied Mathematics and Informatics. Works of the Faculty of Computational Mathematics and Cybernetics of Lomonosov Moscow State University. — 2014. — No. 46. — Pp. 103-127.

Authors' Information

Alexander Dokukin *— Computing Centre of Russian Academy of Sciences, researcher, 40 Vavilova St., Moscow, Russia, 119333; e-mail: dalex@ccas.ru.*
Major Fields of Scientific Research: Algebraic Approach to Pattern Recognition.